



# STIC Search Report

## Biotech-Chem Library

STIC Database Tracking Number: 950345

**TO:** Jeanine Goldberg  
**Location:** mail 12E12; room 12D11  
**Art Unit:** 1634  
**Tuesday, May 27, 2003**

**Case Serial Number:** 09/935464

**From:** Barb O'Bryen  
**Location:** Biotech-Chem Library  
**CM1-6A05**  
**Phone:** 308-4291

*BOB*

**barbara.obryen@uspto.gov**

Seal

**O'Bryen, Barbara**

**From:** Goldberg, Jeanine  
**Sent:** Tuesday, May 20, 2003 6:14 AM  
**To:** O'Bryen, Barbara  
**Subject:** RE: 09/935,464- schizophrenia

Thank you. I called applicant.

1. Please search SEQ ID NO: 12-13 and 39 only.

THanks

-----Original Message-----

**From:** O'Bryen, Barbara  
**Sent:** Monday, May 19, 2003 4:42 PM  
**To:** Goldberg, Jeanine  
**Subject:** RE: 09/935,464- schizophrenia

Hi Jeanine,  
this case has only 90 seqs. How would you like to modify this request?  
Barb

-----Original Message-----

**From:** Goldberg, Jeanine  
**Sent:** Monday, May 19, 2003 3:36 PM  
**To:** O'Bryen, Barbara  
**Subject:** 09/935,464- schizophrenia

1. please search SEQ ID NO: 39, 12-13, 89-92, 100, 101.

THANK YOU  
Jeanine

Jeanine Enewold Goldberg  
1634  
CM1--12D11  
Mailbox-- 12E12  
306-5817

Copyright (c) 1993 - 2003 Compugen Ltd.

Gencore version 5.1.4\_p5\_4578

### OM nucleic - nucleic search, using sw model

Run on: May 24, 2003, 09:39:57 ; Search time 705.6 seconds

(without alignments)  
866.155 Million cell updates/sec

Title: US-09-935-464-13

Perfect score: 21

Sequence: actttccttggcagctgttc 21

Scoring table: IDENTITY.NUC

Gapop 10.0 , gapext 1.0

Searched: 2054610 seqs, 1451402878 residues

Total number of hits satisfying chosen parameters:

4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBank,\*

1: gb\_ba:\*

2: gb\_ntg:\*

3: gb\_in:\*

4: gb\_on:\*

5: gb\_cr:\*

6: gb\_pot:\*

7: gb\_di:\*

8: gb\_pl:\*

9: qb\_p:\*

10: qb\_r:\*

11: qb\_sts:\*

12: qb\_sy:\*

13: qb\_un:\*

14: qb\_vl:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_on:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_vl:\*

29: em\_un:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htg\_mus:\*

34: em\_htg\_pln:\*

35: em\_htg\_rod:\*

36: em\_htg\_mam:\*

37: em\_htg\_vrt:\*

38: em\_htg\_syr:\*

39: em\_htg\_hum:\*

40: em\_htg\_mus:\*

41: em\_htg\_oother:\*

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	21	100.0	157875	9 HS272116	AL023754 Human DNA
c 2	19	90.5	192169	2 AC020849	AC020849 Mus muscu
c 3	19	90.5	206894	2 AC113958	AC113958 Mus muscu
c 4	18.4	87.5	16305	2 AC126986	AC126986 Rattus no
c 5	18.4	87.5	19238	2 AC103908	AC103908 Canis fam
c 6	18.4	87.5	198489	2 AC100820	AC100820 Homo sapi
c 7	18.4	87.6	201000	9 AC105150	AC105150 Homo sapi
c 8	18.4	87.6	201973	2 AC017049	AC017049 Homo sapi
c 9	18.4	87.5	218992	5 AC10903	AC103903 Canis fam
c 10	17.8	84.8	425	5 APERONA	X57379 A. platyrhyn
c 11	17.8	84.8	1871	9 BC005383	BC005383 Homo sapi
c 12	17.8	84.8	1923	9 AB110957	AF110957 Homo sapi
c 13	17.8	84.8	1960	9 AB005969	AB005969 Homo sapi
c 14	17.8	84.8	2050	9 AB002135	AB002135 Homo sapi
c 15	17.8	84.8	2081	9 BC003171	BC003171 Homo sapi
c 16	17.8	84.8	2108	9 BC004129	BC004129 Homo sapi
c 17	17.8	84.8	2582	9 HSM802492	AL157437 Homo sapi
c 18	17.8	84.8	2617	9 HS801873	U85832 Human anthr
c 19	17.8	84.8	2627	9 HS801873	AL136905 Homo sapi
c 20	17.8	84.8	2648	9 AF090384	AF090384 Homo sapi
c 21	17.8	84.8	2660	9 BC003153	BC003153 Homo sapi
c 22	17.8	84.8	2663	9 AK023120	AK023120 Homo sapi
c 23	17.8	84.8	2683	9 AF079566	AF079566 Homo sapi
c 24	17.8	84.8	4097	2 AB002137	AB002137 Homo sapi
c 25	17.8	84.8	5026	5 AB089503	AB089503 Nippomia
c 26	17.8	84.8	11464	3 DROLAMB2	M59417 Drosophila
c 27	17.8	84.8	48836	2 AC020151	AC020151 Drosophil
c 28	17.8	84.8	61213	2 AC103017	AC103017 Homo sapi
c 29	17.8	84.8	6265	2 AC10806	AC10806 Mus muscu
c 30	17.8	84.8	6295	2 AC130171	AC130171 Rattus no
c 31	17.8	84.8	87196	2 AC125497	AC125497 Gallus ga
c 32	17.8	84.8	104848	2 AC103132	AC103132 Rattus no
c 33	17.8	84.8	144749	2 AC025912	AC025912 Mus muscu
c 34	17.8	84.8	174817	2 AC127186	AC127186 Rattus no
c 35	17.8	84.8	17886	2 AC116832	AC116832 Mus muscu
c 36	17.8	84.8	181063	3 AC010107	AC010107 Drosophil
c 37	17.8	84.8	184557	3 AC010403	AC010403 Drosophil
c 38	17.8	84.8	189495	10 AL607083	AL607083 Mouse DNA
c 39	17.8	84.8	194531	2 AC124924	AC124924 Rattus no
c 40	17.8	84.8	285578	3 AE003551	AE003551 Drosophil
c 41	17.4	82.9	43683	2 AC004396	AC004396 Pseudomon
c 42	17.4	82.9	61245	2 AC044824	AC044824 Homo sapi
c 43	17.4	82.9	92822	8 AC005917	AC005917 Arabidops
c 44	17.4	82.9	96102	2 AC11304	AC11304 Rattus no
c 45	17.4	82.9	97495	8 AC003058	AC003058 Arabidops

### ALIGNMENTS

Pred. No. is the number of results predicted by chance to have a

Human DNA sequence from clone 272116 on chromosome 10p2.1-3.3.

Contains the 3' end of the LAMB3 gene for Laminin, Beta 3 (Nicein, Kalinin, BM600) and a novel Rat Ca2+-Calmodulin Dependent Protein Kinase gene. Contains ESTs, STSs, GSS, genomic marker DIS491 and a ca repeat polymorphism, complete sequence.

ACCESSION AI023754

VERSION AI023754.1

DEFINITION Human DNA sequence

KEYWORDS Human DNA sequence

Protein Kinase; DIS491; Kalinin; LAMB3; Laminin Beta 3; Nicein.

SOURCE Homo sapiens.

ORGANISM Homo sapiens.

Bikarrotqa, Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;

REFERENCE 1 (bases 1 to 157875)  
 AUTHORS Graham, D.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-Nov-1998) Sanger Centre, Hinxton, Cambridgeshire,  
 CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk Clone  
 COMMENT requests: clone\_requests@sanger.ac.uk  
 On Dec 12, 1998 this sequence version replaced gi|3873472.  
 During sequence assembly data is compared from overlapping clones.  
 Where differences are found these are annotated as variations  
 together with a note of the overlapping clone name. Note that the  
 variation annotation may not be found in the sequence submission  
 corresponding to the overlapping clone, as we submit sequences with  
 only a small overlap as described above.  
 This sequence is the entire insert of clone 272L16. This sequence  
 has been finished according to sequence map criteria as follows. An  
 attempt is made to resolve all sequencing problems, such as  
 compressions and repeats, but not necessarily within known  
 annotated human repeat sequence elements (e.g. Alu). Where the  
 sequence is ambiguous, there is an annotation using the 'unsure'  
 feature key.  
 272L16 is from the library RPCL1 constructed at the Roswell Park  
 Cancer Institute by the group of Pieter de Jong. For further  
 details see <http://bacpac.med.buffalo.edu/> VECTOR: PCIPAC2 This  
 sequence was generated from a human chromosome 1 bacterial clone  
 1 mapping group and Brian Schutte, Brian Bjork, Kevin Coppage and  
 Jeffrey Murray, Department of Pediatrics, University of Iowa, USA.  
 Further information can be found at  
<http://www.sanger.ac.uk/HGP/Chrl>.

FEATURES

source	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
	(bases 1 to 157875)
FEATURES	
source	

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="q32.1-32.3"  
 /clone\_id="RPCL1-272L16"  
 /clone\_lb="RPCL1-1"  
 966..1078  
 /note="MTR1 repeat: matches 281..410 of consensus"  
 1082..1106  
 /note="MTR1 repeat: matches 1..426 of consensus"  
 repeat\_region  
 1585..1875  
 /note="MTR1 repeat: matches 1..319 of consensus"  
 repeat\_region  
 1898..2024  
 /note="MTR1 repeat: matches 58..203 of consensus"  
 repeat\_region  
 2347..2325  
 /note="MTR34 repeat: matches 6..543 of consensus"  
 repeat\_region  
 3090..3417  
 /note="MTR1A2 repeat: matches 23..374 of consensus"  
 repeat\_region  
 3677..4074  
 /note="MTR54B repeat: matches 486..902 of consensus"  
 repeat\_region  
 4089..5338  
 /note="MTR1A2-internal repeat: matches 358..1543 of  
 consensus"  
 5339..5532  
 /note="AluJo repeat: matches 101..302 of consensus"  
 5536..5666  
 /note="MTR1-internal repeat: matches 5..450 of  
 consensus"  
 5972..6019  
 repeat\_region  
 6020..6159  
 /note="MTR7 repeat: matches 1..450 of consensus"  
 repeat\_region  
 6460..6781  
 /note="MTR1-repeat: matches 1..318 of consensus"  
 repeat\_region  
 6786..6970  
 /note="MTR repeat: matches 56..142 of consensus"  
 repeat\_region  
 8020..8067  
 /note="MTR repeat: matches 95..146 of consensus"  
 8150..8293  
 /note="MTR repeat: matches 58..212 of consensus"  
 repeat\_region  
 8380..9083  
 /note="MTR repeat: matches 2..21 copies 2 mer ca 100% conserved"

repeat\_region  
 /note="MERS50 repeat: matches 1..711 of consensus"  
 9084..9303  
 /note="MIR repeat: matches 13..262 of consensus"  
 repeat\_region  
 9493..9795  
 /note="AluX repeat: matches 1..304 of consensus"  
 repeat\_region  
 10108..10882  
 /note="L2 repeat: matches 2506..2695 of consensus"  
 10829..11005  
 /note="L2 repeat: matches 2671..2748 of consensus"  
 repeat\_region  
 12092..12129  
 /note="19 copies 2 mer ca 84% conserved"  
 12095..12130  
 /note="9 copies 4 mer acac 86% conserved"  
 12172..12855  
 /note="MERNA repeat: matches 20..186 of consensus"  
 repeat\_region  
 13108..13221  
 /note="MIR repeat: matches 80..191 of consensus"  
 14495..15046  
 /note="L2 repeat: matches 2127..2750 of consensus"  
 16671..16859  
 /note="MERA repeat: matches 2..189 of consensus"  
 repeat\_region  
 17103..17476  
 /note="L2 repeat: matches 112..249 of consensus"  
 17679..17938  
 /note="L2 repeat: matches 17..422 of consensus"  
 repeat\_region  
 17891..17953  
 /note="L2 repeat: matches 2694..2749 of consensus"  
 repeat\_region  
 18850..18758  
 /note="L2 repeat: matches 2641..2750 of consensus"  
 repeat\_region  
 19046..1906  
 /note="MERS4 repeat: matches 4..189 of consensus"  
 repeat\_region  
 19460..19871  
 /note="L2 repeat: matches 2309..2748 of consensus"  
 repeat\_region  
 19916..19987  
 /note="MIR repeat: matches 63..140 of consensus"  
 repeat\_region  
 20200..20228  
 /note="MERS4 repeat: matches 510..538 of consensus"  
 repeat\_region  
 20233..20365  
 /note="MERS4 repeat: matches 413..543 of consensus"  
 repeat\_region  
 20338..20388  
 /note="LTR29 repeat: matches 454..503 of consensus"  
 repeat\_region  
 20459..20809  
 /note="MERS4 repeat: matches 2..366 of consensus"  
 repeat\_region  
 2076..21143  
 /note="MERS4 repeat: matches 6..172 of consensus"  
 repeat\_region  
 21834..22278  
 /note="MLTC repeat: matches 1..466 of consensus"  
 repeat\_region  
 22477..22528  
 /note="25 copies 2 mer ag 79% conserved"  
 repeat\_region  
 22826..22937  
 /note="MERL repeat: matches 1..112 of consensus"  
 repeat\_region  
 24497..25034  
 /note="MERS3 repeat: matches 30..171 of consensus"  
 repeat\_region  
 25084..25593  
 /note="MIR repeat: matches 26..145 of consensus"  
 repeat\_region  
 25207..25335  
 /note="MERL repeat: matches 13..259 of consensus"  
 repeat\_region  
 25919..26633  
 /note="MERS3 repeat: matches 68..541 of consensus"  
 repeat\_region  
 26564..26813  
 /note="MERA2 repeat: matches 1..503 of consensus"  
 repeat\_region  
 26814..26885  
 /note="MTR repeat: matches 1..68 of consensus"  
 repeat\_region  
 26831..27357  
 /note="L2 repeat: matches 2292..2723 of consensus"  
 repeat\_region  
 27779..27820  
 /note="21 copies 2 mer ca 100% conserved"

repeat\_1<sup>repe-</sup>  
repeat\_1<sup>repe-</sup>

repeat\_region 2779..27818  
 /note="10 copies 4 mer caca 100% conserved"  
 repeat\_region 29036..29195  
 /note="HML repeat: matches 4..163 of consensus"  
 repeat\_region 29263..29500  
 /note="HML repeat: matches 141..355 of consensus"  
 repeat\_region 29501..29820  
 /note="MITAL repeat: matches 1..365 of consensus"  
 repeat\_region 29821..30076  
 /note="HML repeat: matches 35..618 of consensus"  
 repeat\_region 30075..30248  
 /note="TRAM A repeat: matches 7..142 of consensus"  
 repeat\_region 30246..30440  
 /note="HML repeat: matches 683..874 of consensus"  
 repeat\_region 30339..30788  
 /note="Charliela repeat: matches 1189..1455 of consensus"  
 repeat\_region 30789..31099  
 /note="AluG repeat: matches 1..306 of consensus"  
 repeat\_region 31100..31167  
 /note="Charliela repeat: matches 1..125..1189 of consensus"  
 repeat\_region 31170..31914  
 /note="L1M2 repeat: matches 5448..6190 of consensus"  
 repeat\_region 31915..32141  
 /note="L1M2 repeat: matches 4317..4530 of consensus"  
 repeat\_region 32142..32351  
 /note="MSTA repeat: matches 5..426 of consensus"  
 repeat\_region 32352..32356  
 /note="L1M2 repeat: matches 3451..4317 of consensus"  
 repeat\_region 33437..33756  
 /note="AluY repeat: matches 1..309 of consensus"  
 repeat\_region 3357..3401  
 /note="L1M2 repeat: matches 761..3451 of consensus"  
 repeat\_region 37817..38850  
 /note="Charliela repeat: matches 1..1142 of consensus"  
 repeat\_region 39233..39437  
 /note="L1M2 repeat: matches 1..140 of consensus"  
 misc\_feature complement(3954..39839)  
 /note="match: GSS A0070531"  
 repeat\_region 41165..41445  
 /note="L2 repeat: matches 2180..2489 of consensus"  
 repeat\_region 41615..41810

Query Match 100% Score 21; DB 9; Length 157875;  
 Best Local Similarity 100.0%; Pread: No. 6.1; Matches 21; Conservatve 0; Mismatches 0; Indels 0; Gaps 0;

Qr 1 ACTTCCTTGGACGCTGTGTC 21

Db 143687 ACTTCTCTGGACGCTGTGTC 143657

---

RESULT 2

AC020849/C AC020849 192169 bp DNA LINEAR HTG 15-JUL-2000  
 LOCUS AC020849 Mus musculus clone RP21-43903; WORKING DRAFT SEQUENCE; 55 unordered pieces.  
 ACCESSION AC020849  
 VERSION AC020849\_4 GI:9211211  
 KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
 SOURCE Mus musculus.  
 ORGANISM Mus musculus.  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi; Mammalia; Eutheria; Rodentia; Sciuromorpha; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 192169)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Sequencing of Mouse  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 192169)\*  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 COMMENT On Jul 15, 2000 this sequence version replaced gi:6930195.

-----Genome Center  
 Center: Joint Genome Institute  
 Center Code: JGI  
 Web site: http://www.jgi.doe.gov

Project Information  
 Center Project Name: 1426557  
 Center clone name: RPCI-21\_43909

Summary Statistics

Consensus quality: 149566 bases at least 940  
 Consensus quality: 170600 bases at least 930  
 Consensus quality: 175179 bases at least 920  
 Estimated insert size: 180000; pulse field gel estimation  
 Quality coverage: 3..49 in 020 bases; pulse field gel estimation  
 Quality coverage: 3..37 in 020 bases; sum of contigs estimation  
 \* NOTE: This is a 'working draft' sequence. It currently  
 \* consists of 35 contigs. The true order of the pieces  
 \* is not known and their order in this sequence record is  
 \* arbitrary. Gaps between the contigs are represented as  
 \* runs of N, but the exact sizes of the gaps are unknown.  
 \* This record will be updated with the finished sequence  
 \* as soon as it is available and the accession number will  
 \* be preserved.

*	1236:	contig of 1236 bp in length
*	1237:	1356: gap of unknown length
*	1337:	2710: contig of 1374 bp in length
*	2810:	2810: gap of unknown length
*	2811:	4014: contig of 1204 bp in length
*	4015:	4114: gap of unknown length
*	4115:	5404: contig of 1290 bp in length
*	5405:	5503: gap of unknown length
*	5505:	7236: contig of 1752 bp in length
*	7257:	736: gap of unknown length
*	7357:	8829: contig of 1473 bp in length
*	8830:	8929: gap of unknown length
*	8930:	10243: contig of 1312 bp in length
*	10242:	10311: gap of unknown length
*	10342:	12012: contig of 1671 bp in length
*	12013:	12112: gap of unknown length
*	12113:	13606: contig of 1494 bp in length
*	13607:	13706: gap of unknown length
*	13702:	15052: contig of 1346 bp in length
*	15053:	15152: gap of unknown length
*	15153:	16157: contig of 1005 bp in length
*	16158:	16257: gap of unknown length
*	16258:	17500: contig of 1323 bp in length
*	17501:	17581: gap of unknown length
*	17581:	17681: contig of 1320 bp in length
*	17681:	19000: gap of unknown length
*	19001:	19100: gap of unknown length
*	19101:	20952: contig of 1852 bp in length
*	20953:	21052: gap of unknown length
*	21053:	22865: contig of 1814 bp in length
*	22865:	22865: gap of unknown length
*	22867:	22967: contig of 2304 bp in length
*	22967:	25220: gap of unknown length
*	25221:	25370: contig of 1437 bp in length
*	25371:	26807: gap of unknown length
*	26808:	26907: gap of unknown length
*	26908:	28347: contig of 1440 bp in length
*	28347:	28447: gap of unknown length
*	28448:	28448: contig of 2011 bp in length
*	30459:	30558: gap of unknown length
*	30559:	31680: contig of 1128 bp in length
*	31687:	3176: gap of unknown length
*	3176:	31787: contig of 1452 bp in length
*	31787:	3228: contig of 3330 bp in length
*	3228:	3239: gap of unknown length
*	3239:	3330: contig of 1980 bp in length
*	3330:	35318: gap of unknown length
*	35319:	35418: contig of 1567 bp in length
*	35419:	36985: gap of unknown length
*	36985:	37085: contig of 2076 bp in length
*	37086:	39161: gap of unknown length

Copyright (c) 1993 - 2003 Compugen Ltd.

### OM nucleic - nucleic search, using sw model

Run on: May 24, 2003, 09:39:57 ; Search time 604.8 Seconds  
                   (without alignments)  
                   866.155 Million cell updates/sec

Title: JS-09-935-464-12  
       Sequence: 1 tggagcttggggaga 18

Scoring table: IDENTITY\_NUC  
       Gapop 10.0 , Gapext 1.0

Searched: 205460 seqs, 1455140878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
                   Maximum Match 100%

Listing first 45 summaries

Database : GenEmbl:\*

1: gb\_ba:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_pat:\*

7: gb\_ph:\*

8: gb\_pl:\*

9: gb\_pr:\*

10: gb\_to:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_ba:\*

16: em\_fan:\*

17: em\_hun:\*

18: em\_in:\*

19: em\_mn:\*

20: em\_on:\*

21: em\_oi:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_pi:\*

26: em\_ro:\*

27: em\_sis:\*

28: em\_vl:\*

29: em\_vn:\*

30: em\_htg\_hum:\*

31: em\_htg\_inv:\*

32: em\_htg\_other:\*

33: em\_htg\_lms:\*

34: em\_htg\_p1n:\*

35: em\_htg\_r1d:\*

36: em\_htg\_m1m:\*

37: em\_htg\_vrt:\*

38: em\_sv:\*

39: em\_hugo\_num:\*

40: em\_hugo\_mus:\*

41: em\_hugo\_other:\*

score greater than or equal to the score of the result being printed,  
                   and is derived by analysis of the total score distribution.

### SUMMARIES

Result No.	Score	Query	% Match	Length	DB ID	Description
1	18	100.0	157875	9	HS272L16	AL023754 Human DNA
2	18	100.0	215293	2	AC122914	AC122914 Mus muscu
3	17	94.4	78508	2	AC018191	AC018191 Drosophil
4	17	94.4	139924	2	AC11512	AC11512 Mus muscu
5	17	94.4	144723	10	AC005118	AC005118 Mus muscu
6	17	94.4	154639	2	AC090175	AC090175 Homo sapi
7	17	94.4	158458	9	AC007395	AC007395 Homo sapi
8	17	94.4	159836	9	AL292085	AL292085 Human DNA
9	17	94.4	163597	2	AL162589	AL162589 Homo sapi
10	17	94.4	171569	3	AC007811	AC007811 Drosophil
11	17	94.4	172686	2	AC128905	AC128905 Rattus no
12	17	94.4	174215	2	AC094823	AC094823 Rattus no
13	17	94.4	175420	2	AC129161	AC129161 Rattus no
14	17	94.4	176735	3	AC007810	AC007810 Drosophil
15	17	94.4	177095	2	AC113177	AC113177 Mus muscu
16	17	94.4	177451	2	AC009154	AC009154 Rattus no
17	17	94.4	183057	2	AC117132	AC117132 Rattus no
18	17	94.4	186057	2	AC123412	AC123412 Rattus no
19	17	94.4	186284	2	AC009153	AC009153 Rattus no
20	17	94.4	196785	2	AC120065	AC120065 Rattus no
21	17	94.4	197768	2	AL044174	AL044174 Mus muscu
22	17	94.4	204653	10	AC005102	AC005102 Mus muscu
23	17	94.4	210133	9	AC013564	AC013564 Homo sapi
24	17	94.4	218391	3	AE003719	AE003719 Drosophil
25	17	94.4	226919	2	AL845275	AL845275 Mus muscu
26	16.4	91.1	1021	10	AB036008	AB036008 Human DNA
27	16.4	91.1	5229	9	AL9804796	AL833483 Homo sapi
28	16.4	91.1	5543	10	AF175432	AF175432 Mus muscu
29	16.4	91.1	5877	10	AF036009	AF036009 Mus muscu
30	16.4	91.1	6000	10	AL57410502	AF154111 Mus muscu
31	16.4	91.1	9636	10	AB056445	AB056445 Mus muscu
32	16.4	91.1	2111	9	AL990116	AL990116 Human DNA
33	16.4	91.1	77951	2	AC116475	Continuation (6 of AC106419 Rattus no
34	16.4	91.1	80146	2	AC106419	AC106419 Rattus no
35	16.4	91.1	84307	2	AC023206	AC023206 Homo sapi
36	16.4	91.1	88557	2	AL54986	AL54986 Homo sapi
37	16.4	91.1	104441	2	AC095721	AC095721 Rattus no
38	16.4	91.1	106639	2	AC121371	AC121371 Rattus no
39	16.4	91.1	107171	2	AP004152	AP004152 Oryza sat
40	16.4	91.1	112392	9	AI357500	AI357500 Human DNA
41	16.4	91.1	128332	9	AC068657	AC068657 Homo sapi
42	16.4	91.1	128517	2	AC094352	AC094352 Rattus no
43	16.4	91.1	131105	9	AC008000	AC008000 Homo sapi
44	16.4	91.1	137849	2	AC105141	AC105141 Homo sapi
45	16.4	91.1	141399	2	AC110221	AC110221 Mus muscu

### ALIGNMENTS

RESULT 1  
       HS272L16  
       LOCUS  
       Human DNA sequence from clone 272L16 on chromosome 19q2.1-2.3.  
       Contains the 3' end of the LAMB3 gene for Laminin, Beta 3 (Nicein, Kalinin, BM600) and a novel Rat Ca2+-Calmodulin dependent Protein Kinase like gene. Contains ESTs, SRS, GSS, genomic marker Dis491 and a ca repeat polymorphism, complete sequence.  
       ACCESSION AL023754  
       VERSION AL023754.1  
       KEYWORDS HTG; BM600; ca repeat polymorphism; Ca2+/Calmodulin dependent Protein Kinase; Dis491; Kalinin; LAMB3; Laminin Beta 3; Nicein.  
       SOURCE  
       ORGANISM Homo sapiens  
       Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Pred. No. is the number of results predicted by chance to have a

REFERENCE 1 (bases 1 to 15875)  
 AUTHORS Graham,D.  
 TITLE Direct Submission  
 JOURNAL Submitted (27-Nov-1998) Sanger Centre, Hinxton, Cambridgeshire, CB10 1EA, UK. E-mail enquiries: humg@y.sanger.ac.uk Clone requests: clone@y.sanger.ac.uk

COMMENT On Dec 12, 1998 this sequence version replaced gi|387472. During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.

This sequence is the entire insert of clone 272L16. This sequence has been finished according to sequence map criteria as follows. An attempt is made to resolve all sequencing problems such as compressions and repeats, but not necessarily within known annotated human repeat elements (e.g. Alu). Where the sequence is ambiguous, there is an annotation using the 'unsure' feature key.

272L16 is from the library RPCII constructed at the Roswell Park Cancer Institute by the group of Pieter de Jong. For further details see <http://hgac.med.buffalo.edu/VECTOR:PCPAC2>. This sequence was generated from a human chromosome 1 bacterial clone contig constructed in collaboration with the Sanger Centre chromosome 1 mapping group and Brian Schutte, Brian Bjork, Kevin Coppage and Jeffrey Murray, Department of Pediatrics, University of Iowa, USA. Further information can be found at <http://www.sanger.ac.uk/HGP/Chrl>.

FEATURES

source

1. Location/Qualifiers

/organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="1"  
 /map="q32.1-32.3"  
 /clone="RP1-272L16"  
 /clone\_id="RPCII-1"  
 966..1078

/note="MIR repeat: matches 281..410 of consensus"  
 1082..1506

/note="MSTR repeat: matches 1..426 of consensus"  
 1585..1875

/note="MTRAL repeat: matches 1..319 of consensus"  
 1898..2024

/note="MTR repeat: matches 58..203 of consensus"  
 2347..2925

/note="MTR34 repeat: matches 5..543 of consensus"  
 3090..3417

/note="MTR12 repeat: matches 23..374 of consensus"  
 3677..4074

/note="MERS54B repeat: matches 486..902 of consensus"  
 4083..5338

/note="MTR12-internal repeat: matches 358..1643 of consensus"  
 5339..5532

/note="Alu repeat: matches 101..302 of consensus"  
 5336..5566

/note="MTRAL-internal repeat: matches 5..450 of consensus"  
 5972..6019

/note="MTRAL repeat: matches 318..365 of consensus"  
 6030..6459

/note="LTR repeat: matches 1..450 of consensus"  
 6460..6781

/note="MTRAL repeat: matches 1..318 of consensus"  
 6786..6878

/note="MTR repeat: matches 56..142 of consensus"  
 8030..8067

/note="MTR repeat: matches 95..146 of consensus"  
 8150..8293

/note="MIR repeat: matches 68..212 of consensus"  
 8300..9083

repeat\_region

9084..9033

/note="MIR repeat: matches 13..262 of consensus"  
 9433..9795

/note="Alus repeat: matches 1..304 of consensus"  
 10108..10282

repeat\_region

10929..11005

/note="L2 repeat: matches 2656..2695 of consensus"  
 12092..12129

/note="19 copies 2 mer ca 84% conserved"  
 12095..12130

repeat\_region

12722..12855

/note="MERNA repeat: matches 20..186 of consensus"  
 13108..13221

/note="MIR repeat: matches 80..191 of consensus"  
 14495..15046

repeat\_region

14235..14279

/note="MIR repeat: matches 60..105 of consensus"  
 14427..14442

repeat\_region

16671..16659

/note="MERNA repeat: matches 2..189 of consensus"  
 17103..17476

repeat\_region

17679..17938

/note="L2 repeat: matches 17..422 of consensus"  
 17891..17953

repeat\_region

18050..18158

/note="L2 repeat: matches 2127..2750 of consensus"  
 18246..18906

repeat\_region

19160..19171

/note="MERNA repeat: matches 7..262 of consensus"  
 19316..19387

repeat\_region

20200..20228

/note="MER repeat: matches 63..140 of consensus"  
 20233..20365

repeat\_region

20338..20388

/note="MERS4 repeat: matches 510..538 of consensus"  
 21334..22278

repeat\_region

20359..20509

/note="MERS4 repeat: matches 63..140 of consensus"  
 22477..22528

repeat\_region

22836..22937

/note="MER4A repeat: matches 2..356 of consensus"  
 23095..23143

repeat\_region

23134..23188

/note="MTR34 repeat: matches 6..172 of consensus"  
 25084..25193

repeat\_region

25207..25435

/note="MER81 repeat: matches 1..112 of consensus"  
 24897..25034

repeat\_region

25919..26333

/note="MERS5A repeat: matches 30..171 of consensus"  
 25304..25613

repeat\_region

26814..26885

/note="MTR repeat: matches 13..259 of consensus"  
 26931..27357

repeat\_region

27779..27820

/note="L2 repeat: matches 2292..2723 of consensus"  
 27779..27820

repeat\_region

/note="21 copies 2 mer ca 100% conserved"

May 27 7

repeat\_region 2779..27818 /note="10 copies 4 mer caca 10% conserved"

repeat\_region 29035..29195 /note="Hail repeat: matches 4..163 of consensus"

repeat\_region 29263..29500 /note="Hail repeat: matches 141..355 of consensus"

repeat\_region 29501..29820 /note="MM1AL repeat: matches 1..365 of consensus"

repeat\_region 29821..30076 /note="Hail repeat: matches 355..618 of consensus"

repeat\_region 30095..30228 /note="FLAM\_A repeat: matches 7..142 of consensus"

repeat\_region 30246..30440 /note="Hail repeat: matches 683..874 of consensus"

repeat\_region 30539..30788 /note="Charliela repeat: matches 1189..1455 of consensus"

repeat\_region 30789..31099 /note="AluG repeat: matches 1..306 of consensus"

repeat\_region 31100..31167 /note="Charliela repeat: matches 1125..1189 of consensus"

repeat\_region 31170..31914 /note="LIMA4 repeat: matches 5448..6190 of consensus"

repeat\_region 31923..32141 /note="LIM2 repeat: matches 5..426 of consensus"

repeat\_region 32142..32251 /note="MM1AL repeat: matches 1..140 of consensus"

repeat\_region 32252..33436 /note="LIM2 repeat: matches 3451..4317 of consensus"

repeat\_region 33437..33756 /note="AluY repeat: matches 1..309 of consensus"

repeat\_region 33757..33604 /note="LIM2 repeat: matches 761..3451 of consensus"

repeat\_region 37817..37850 /note="Charliela repeat: matches 1..1142 of consensus"

repeat\_region 39293..39437 /note="MR repeat: matches 1..140 of consensus"

repeat\_region 41165..41445 /note="L2 repeat: matches 2180..2489 of consensus"

repeat\_region 41615..41810 /note="match: GSS A0070531"

Query Match 100%; Score 18; DB 9; Length 157875;

Best Local Similarity 100.0%; Pred. No. 29;

Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 TGGAGCTTGGGGAGCA 18

D<sub>b</sub> 143358 TGGAGCTTGGGGAGCA 143375

RESULT 2

AC122914 LOCUS AC122914 215293 bp DNA LINEAR HGNC 28-MAY-2002

DEFINITION Mus musculus chromosome 10K clone RP3-26F9, WORKING DRAFT

SEQUENCE Mus musculus chromosome 10K clone RP3-26F9, WORKING DRAFT

ACCESSION AC122914

VERSION AC122914.1 GI:21218545

KEYWORDS HG, HGs\_PHASE1, HGs\_DRAFT.

SOURCE Mus musculus.

ORGANISM Mus musculus.

Bukarica; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Buteraria; Rodentia; Sciurognathai; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 215293)

AUTHORS McPherson,J.D. and Waterston,R.H.

TITLE The sequence of *Mus musculus* clone

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 215293)

AUTHORS McPherson,J.D. and Waterston,R.H.

TITLE Direct Submission

JOURNAL Submitted (28-MAY-2002) Genome Sequencing Center, 4444 Forest Park Parkway, St. Louis, MO 63108, USA

COMMENT

----- Genome Center -----

Center: Washington University Genome Sequencing Center

Center code: WUGSC

Web site: <http://genome.wustl.edu/gsc/index.shtml>

Contact: submissions@watson.wustl.edu

----- Project Information -----

Project name: M\_BA0026P09

----- Summary Statistics -----

Sequencing vector: M13; 0%

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.990319

Sequencing vector: Plasmid; 100% of reads

Chemistry: Dye-primer ET; 0% of reads

Consensus quality: 211324 bases at least Q30

Consensus quality: 211779 bases at least Q20

Insert size: 212000; agarose-fp

Insert size: 2124293; sum-of-contigs

Quality coverage: 12.87 in Q20 bases; agarose-fp

Quality coverage: 10.50 in Q20 bases; sum-of-contigs

\* NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown.

\* This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

\* \* \* \* \*

\* 1 1233: contig of 1233 bp in length

\* 1234 1333: gap of unknown length

\* 1334 2576: contig of 1243 bp in length

\* 2577 2676: gap of unknown length

\* 2677 4229: contig of 153 bp in length

\* 4230 4329: gap of unknown length

\* 4330 9108: contig of 4779 bp in length

\* 9109 9208: gap of unknown length

\* 9209 15839: contig of 6631 bp in length

\* 15840 15939: gap of unknown length

\* 15940 26130: contig of 10191 bp in length

\* 26131 26230: gap of unknown length

\* 26231 26233: contig of 23813 bp in length

\* 26233 50043: gap of unknown length

\* 50044 50143: gap of unknown length

\* 50144 74268: contig of 24125 bp in length

\* 74269 74368: gap of unknown length

\* 74369 101258: contig of 26890 bp in length

\* 101259 101358: gap of unknown length

\* 101359 142855: contig of 41497 bp in length

\* 142856 142955: gap of unknown length

\* 142956 21523: contig of 72338 bp in length.

FEATURES

Source

1. 215293 /organism="Mus musculus"  
     /DB\_XREF="Taxon:10090"  
     /chromosome="10K"  
     /clone="RP3-26F9"

misc\_feature

1. 1233 /note="assembly\_name:Contig10"

misc\_feature

1334. 2575 /note="assembly\_name:Contig11"

misc\_feature

2677. 4229 /note="assembly\_name:Contig13"

misc\_feature

4330. 9108 /note="assembly\_name:Contig14"

misc\_feature

9209. 15839 /note="assembly\_name:Contig15"

misc\_feature

15940. 26130 /note="assembly\_name:Contig16"

misc\_feature

26231. 50043 /note="assembly\_name:Contig17"

misc\_feature

50144. 74268 /note="assembly\_name:Contig18"

[http://www.tigr.org/tgb/hungen/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tgb/hungen/bac_end_search/bac_end_search.html).

Seq primer: M13 Reverse  
Class: BAC ends.

Location/Qualifiers

1..388

/organism="Homo sapiens"

/ab\_xref="taxon:9606"

/clone="2367C8"

/clone\_lib="CT-HSP"

/sex="Male"

/cell\_type="Sperm"

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

HindIII" vector: pBeloBAC11; Site-1: HindIII; Site-2:

/note="Vector: pBeloBAC11; Site-1: HindIII; Site-2:

BASE COUNT 102 a 88 c 105 g 93 t

ORIGIN

Query Match 100.0%; Score 18; DB 9; Length 445;  
Best Local Similarity 100.0%; Pred. No. 6.2e+02;  
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGAGCTTGGGGAGCA 18

Db 233 TGGGAGCTTGGGGAGCA 250

RESULT 2

LOCUS A215131 445 bp mRNA linear EST 21-OCT-1998

DEFINITION gp41f02.x1 NCI-CGAP\_Co8 Homo sapiens cDNA clone IMAGE-1925595\_3

ACCESSION A215131

VERSION A215131.1 GI:3778732

KEYWORDS EST.

SOURCE human.

ORGANISM Homo sapiens

Mammalia; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Buteraria; Rodentia; Sciurognathi; Muridae; Murinae; Mus;

Zhao,S., Nierman,W., Feldblum,T., Malek,J., Shatsman,S., Akinet,B., Levin,M., McRann,S., Tsedaye,G., Geerk,K., Kroli,M., de Jong,P. and Fraser,C.M.

Mouse BAC End Sequences from Library RPCI-23

Unpublished (1999)

Other\_GSS: RPCI-23-148E23.TV

Contact: Shaying Zhao

Department of Eukaryotic Genomics

The Institute for Genomic Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: shao@tigr.org

Copies are derived from the mouse BAC library RPCI-23. For BAC

library availability, Please contact Pieter de Jong

(pieter@edetong.med.buffalo.edu). Clones may be Purchased from

BAC/PAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)

or from Resea ch Genetics ([info@resegen.com](http://info@resegen.com)). BAC end page:

[http://www.tigr.org/tgb/bac\\_ends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tgb/bac_ends/mouse/bac_end_intro.html)

Plate: 148 row: E column: 23

Seq primer: SP6

Class: BAC ends

Location/Qualifiers

FEATURES Source

1..480

/organism="Mus musculus"

/strain="C57BL/6J"

/db\_xref="taxon:10090"

/clone="RPCI-23-148E23"

/clone\_lib="RPCI-23"

/sex="Female"

/lab\_host="DH10B"

/note="Organ: Kidney; Brain; Vector: pBACe3; Site-1:

EcoRI; Site-2: EcoRI; Female C57BL/6J mouse kidney and/or

brain genomic DNA was isolated and partially digested

with a combination of EcoRI and EcoRI Methylase. Size

selected DNA was cloned into the pBACe3.6 vector at the

EcoRI sites. The ligation products were transformed into

DH10B electrocompetent cells (BRL Life Technologies).

BASE COUNT 96 a 99 g 144 t

ORIGIN

Query Match 94.4%; Score 17; DB 17; Length 480;

Best Local Similarity 100.0%; Pred. No. 1.7e+03;

Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGGAGCTTGGGGAGCA 17

Db 449 TGGGAGCTTGGGGAGC 465

RESULT 4

FEATURES Source

1..388

/organism="Homo sapiens"

/ab\_xref="taxon:9606"

/clone="IMage:1925595"

/clone\_lib="NCI-CGAP\_Co8"

/tissue\_type="adenocarcinoma"

/lab\_host="DH10B"

/note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a

modified polylinker; 1st strand cDNA was prepared from

colon adenocarcinoma, and was then primed with a Not I -

Oligo(dT) primer. Double-stranded cDNA was ligated to Eco

RI adaptors (Pharmacia), digested with Not I and cloned

into the Not I and Eco RI sites of the modified pT7T3

vector. Library is normalized. Library was constructed by

Bento Soares and M. Fatima Bonaldo."

BASE COUNT 108 a 91 c 124 g 121 t 1 others

ORIGIN

39.



[http://www.tigr.org/db/human/bac\\_end\\_search.html](http://www.tigr.org/db/human/bac_end_search.html).

Seq\_Primer: M13 Reverse  
Class: BAC ends.  
FEATURES  
Source  
Location/Qualifiers

1. .388

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="2367C8"  
/clone\_1.lib="CIT-HSP"  
/sex="Male"

/cell\_type="Sperm"  
/note="Vector: pBeloBAC11; Site\_1: HindIII; Site\_2:  
HindIII; HindIII; Site\_2:

BASE COUNT 102 a 88 c 105 g 93 t  
ORIGIN

Query Match 100.0%; Score 21; DB 17; Length 388;  
Best Local Similarity 100.0%; pred. No. 5.1;  
Matches 21; Conservative 0; Mismatches 0;  
Indels 0; Gaps 0;

QY 1 GATACCCCGCTCATGAG 21  
|||||||.....|||||||  
Db 360 GATACCCCGCTCATGAG 380

RESULT 2

BQ086330 LOCUS BQ086330 573 bp mRNA linear EST 29-APR-2002  
DEFINITION 1j2c07.y1 Melton Normalized Human Islet 4 N4-H1S 1 Homo sapiens  
CDNA clone IMAGE:6135181 5', similar to TR:095523 095523 DJ22L15.1  
, mRNA sequence.

ACCESSION BQ086330  
VERSION BQ086330.1  
KEYWORDS EST

SOURCE  
ORGANISM  
Human  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 573)  
AUTHORS Melton,D., Brown,J., Kent,G., Permutt,A., Lee,C., Kaestner,K.,  
Demischka,I., Scarfone,M., Brastelli,J., Gradioli,G., Clifton,S.,  
Hillier,L., Marras,M., Pape,D., Wylie,T., Martin,J., Blustein,A.,  
Schmitt,A., Theising,B., Ritter,E., Rokko,I., Bennett,J., Cardenas,  
, Gibbons,M., McCann,R., Cole,R., Tsagareishvili,R., Williams,T.  
TITLE Endocrin Pancreas Consortium  
JOURNAL Unpublished (2000)  
COMMENT Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1912  
Fax: 617-495-8257  
Email: dmelton@hmc.harvard.edu  
Library was constructed by Dr. Douglas Melton DNA sequencing by:  
Washington University Genome Sequencing Center This clone is  
available royalty-free through LBNL; please contact the IMAGE  
consortium (<http://infoimage.llnl.gov>) for further information  
Seq primer: -40RP from Gibco  
High quality sequence stop: 447.  
Location/Qualifiers

FEATURES  
Source

1. .373

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6135181"  
/clone\_1.lib="CIT-HSP"  
/sex="Both"  
/tissue\_type="Islets of Langerhans"  
/dev\_stage="Adult"  
/lab\_host="DHL0B"  
/note="Organ: Pancreas; Vector: pBeloBAC11; Site\_1: HindIII; Site\_2:  
Sal 1; Starting library constructed using

Superscript Plasmid library kit (Life Technologies). cDNA  
made by oligo-dT priming. Size-selected by column  
fractionation; average insert size 1.08 kb. Library was  
amplified once on solid support and plasmid DNA from  
library was prepared. The library DNA was normalized by  
method #4 from Bonaldo, Lennon, and Soares 1996 Genome  
Research 6:791-806; 0.5 microgram single-stranded library  
Plasmid DNA was mixed with 5 micrograms PCR product  
representing library inserts and hybridized to an pool of  
20. Single-stranded (unhybridized) plasmids were removed  
by hydroxylapatite chromatography and used to make  
tilts  
library.

BASE COUNT 143 a 177 c 136 g 117 t  
ORIGIN

Query Match 100.0%; Score 21; DB 14; Length 573;  
Best Local Similarity 100.0%; pred. No. 5.6;  
Matches 21; Conservative 0; Mismatches 0;  
Indels 0; Gaps 0;

QY 1 GATACCCCGCTCATGAG 21  
|||||||.....|||||||  
Db 149 GATACCCCGCTCATGAG 169

RESULT 3

B1818261 LOCUS B1818261 742 bp mRNA linear EST 04-OCT-2001  
DEFINITION 603032510P1 NIH\_MGC\_115 Homo sapiens cDNA clone IMAGE:5173587 5',  
mRNA sequence.

ACCESSION B1818261  
VERSION B1818261.1  
KEYWORDS EST

SOURCE  
ORGANISM  
Human  
Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 742)  
AUTHORS NIH-MGC <http://nuc.nci.nih.gov/>.  
TITLE National Institutes of Health, Mammalian Gene Collection (MGC).  
JOURNAL Unpublished (1999)  
COMMENT Contact: Robert Strausberg, Ph.D.  
Email: [craps@nig.hih.gov](mailto:craps@nig.hih.gov)  
Tissue Procurement: Life Technologies, Inc.  
CDNA Library Preparation: Life Technologies, Inc.  
DNA Library Arranged by: The I.M.A.G.E. Consortium (LNL)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LNL at:  
<http://Image.LNL.gov>  
Plate: LAM1432 row: d column: 04  
High quality sequence stop: 742.  
Location/Qualifiers

FEATURES  
Source

1. .742

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:5173587"  
/clone\_1.lib="NIH\_MGC\_115"  
/lab\_host="DHL0B"  
/note="Organ: pooled brain, lung, testis; vector:  
PCMV-SPORT6; Site\_1: NotI; Site\_2: EcoRV (destroyed); RNA  
source anonymous pool of 6 male brains, age range 23-27; 1  
male lung, age 27; and 1 male testis, age 69. Library is  
oligo-dT primed and directionally cloned (EcoRV site is  
destroyed upon cloning). Average insert size 1.8 kb,  
insert size range 1-3 kb. Library is normalized and  
enriched for full-length clones and was constructed by  
Gruber (Invitrogen). Research Genetics tracking code  
021. Note: this is a NIH\_MGC Library."

BASE COUNT 197 a 180 c 197 g 168 t  
ORIGIN

Query Match 100.0%; Score 21; DB 13; length 742;  
Best Local Similarity 100.0%; pred. No. 6;

FEATURES  
Source  
Location/Qualifiers

1. .573

/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="IMAGE:6135181"  
/clone\_1.lib="CIT-HSP"  
/sex="Both"  
/tissue\_type="Islets of Langerhans"  
/dev\_stage="Adult"  
/lab\_host="DHL0B"  
/note="Organ: Pancreas; Vector: pBeloBAC11; Site\_1: Not I;  
Site\_2: Sal 1; Starting library constructed using

Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on:

May 24, 2003, 12:18:47 ; search time 1035 seconds

(without alignments)

281.661 Million cell updates/sec

Title: Perfect score: US-09-935-464-12

Sequence: 1 tggagcttggggaga 18

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched:

15154066 seqs.

8037743376 residues

Total number of hits satisfying chosen parameters:

32308132

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:\*

1: em\_estba:\*

2: em\_lestimu:\*

3: em\_lestinu:\*

4: em\_lestmu:\*

5: em\_lestovr:\*

6: em\_lestpli:\*

7: em\_lestro:\*

8: em\_letric:\*

9: qb\_est1:\*

10: qb\_est2:\*

11: qb\_hrc:\*

12: qb\_lest3:\*

13: qb\_lest4:\*

14: qb\_lest5:\*

15: em\_estfun:\*

16: em\_eston:\*

17: qb\_gss:\*

18: em\_gss\_hum:\*

19: em\_gss\_inv:\*

20: em\_gss\_pln:\*

21: em\_gss\_vrt:\*

22: em\_gss\_fun:\*

23: em\_gss\_mam:\*

24: em\_gss\_mus:\*

25: em\_gss\_other:\*

26: em\_gss\_pro:\*

27: em\_gss\_rod:\*

## SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	18	100.0	388 17	AQ077073
2	18	100.0	445 9	A1215131
3	17	94.4	470 17	AZ293753
4	17	94.4	702 17	AZ58794
5	17	94.4	706 17	AZ117160
6	17	94.4	919 12	BG259997

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## ALIGNMENTS

### RESULT 1

AQ077073  
LOCUS CNT-HSP-2367C8-TR CNT-HSP Homo sapiens genomic clone 2367C8, DNA sequence.

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

GSS

human.

Homo sapiens

JOURNAL

COMMENT

Unpublished (1998)

Other GSS: CNT-HSP-2367C8.TF

Contact: Mark Adams

Department of Eukaryotic Genomics

The Institute for Geometric Research

9712 Medical Center Dr., Rockville, MD 20850, USA

Tel: 301 838 0200

Fax: 301 838 0208

Email: mada@georgetown.org

Clones are available from Research Genetics (info@research.com). BAC

end search page:

[http://www.tigr.org/tdb/humden/bac\\_end\\_search/bac\\_end\\_search.html](http://www.tigr.org/tdb/humden/bac_end_search/bac_end_search.html).  
 Seq primer: M13 Reverse  
 Class: BAC ends.  
 Location/Qualifiers

1. .388 "Homo sapiens"  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone\_id="2367C8"  
 /clone\_libr="CIT-HSP"  
 /sex="Male"  
 /cell\_type="Sperm"  
 /note="Vector: pBeloBAC11; Site\_1: HindIII; Site\_2:  
 HindIII"

RESULT 3  
 BASE COUNT 102 a 88 c 105 g 93 t  
 ORIGIN  
 Query Match 100.0%; Score 18; DB 17; Length 388;  
 Best Local Similarity 100.0%; Pred. No. 6.2e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 Best local similarity 100.0%; Pred. No. 6e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TGGAGCTTGGGAGCA 18  
 Db 271 TGGAGCTTGGGAGCA 288

RESULT 2  
 AI215131  
 LOCUS AI215131 NCI\_CGAP\_Co8 Homo sapiens cDNA clone IMAGE:1925953' linear EST 21-OCT-1998  
 DEFINITION similar to TR-008763 008763 PROTEIN KINASE ; mRNA sequence.  
 ACCESSION AI215131  
 VERSION AI215131.1 GI:3778732  
 KEYWORDS EST.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 REFERENCE 1 (bases 1 to 45)  
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/cnicgap.  
 TITLE Tumor Gene Index  
 JOURNAL Unpublished (1997)  
 COMMENT Contact: Robert Strausberg, Ph.D.  
 Email: cgbps1@mail.nih.gov  
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.  
 Emmert-Buck, M.D., Ph.D.  
 CDNA Library Preparation: M. Bento Soares, Ph.D.  
 CDNA Library Arrayed by: Greg Lennon, Ph.D.  
 DNA Sequencing by: Washington University Genome Sequencing Center  
 Clone distribution: NCI-CGAP clone distribution information can be  
 found through the I.M.A.G.E. Consortium/LINL at:  
<http://www.ncbi.nlm.nih.gov/obrp/Image/image.html>  
 Seq primer: -doup from Gibco  
 High quality sequence stop: 444.

FEATURES source  
 1. .445 "Homo sapiens"  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /clone="IMAGE:1925953"  
 /clone\_id="NCI\_CGAP\_Co8"  
 /tissue\_type="adenocarcinoma"  
 /lab\_host="DH10B"  
 /note="Organ: colon; Vector: pTT3D-Pac (Pharmacia) with a  
 modified polylinker; 1st strand cDNA was prepared from  
 colon adenocarcinoma, and was then primed with a Not I -  
 Eco RI primer. Double-stranded cDNA was ligated to Eco  
 RI adaptors (Pharmacia), digested with Not I and cloned  
 into the Not I and Eco RI sites of the modified pR733  
 vector. Library is normalized. Library was constructed by  
 Bento Soares and M. Fatima Bonaldo."  
 BASE COUNT 108 a 91 c 124 g 121 t 1 others.

Query Match 100.0%; Score 18; DB 9; Length 445;  
 Best Local Similarity 100.0%; Pred. No. 6.2e+02;  
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 TGGAGCTTGGGAGCA 18  
 Db 449 TGGAGCTTGGGAGC 465

RESULT 3  
 LOCUS A2293753  
 DEFINITION RPCI-23-148E23.TU RPCI-23 Mus musculus genomic clone RPCI-23-148E23  
 ACCESSION A2293753  
 VERSION A2293753.1 GI:9535627  
 AUTHORS B., Levinson,M., Mcginn,S., Tsagkave,G., Geer,K., Krolik,M., de Jong,P.  
 ORGANISM Mus musculus  
 Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 REFERENCE 1 (bases 1 to 480)  
 Zhao,S., Nieman,W., Felloblyum,T., Malek,J., Shatsman,S., Akirrett,  
 B., Levinson,M., Mcginn,S., Tsagkave,G., Geer,K., Krolik,M., de Jong,P.  
 and Fraser,C.M.  
 TITLE Mouse BAC End Sequences from Library RPCI-23  
 JOURNAL Unpublished (1999)  
 COMMENT Other\_GSS: RPCI-23-148E23.TU  
 Contact: Shaving Zhao  
 Department of Eukaryotic Genomics  
 9712 Medical Center Dr., Rockville, MD 20850, USA  
 Tel: 301 838 0200  
 Fax: 301 838 0208  
 Email: szhao@tigr.org  
 Clones are derived from the mouse BAC library RPCI-23. For BAC  
 library availability, please contact Pieter de Jong  
<http://pieterdejong.med.buffalo.edu>. Clones may be purchased from  
 BAC/PAC Resources (<http://bacpac.med.buffalo.edu/orderframe.htm>)  
 or from Resea ch Genetics (<http://inforesgen.com>). BAC end page:  
[http://www.tigr.org/tdb/bacends/mouse/bac\\_end\\_intro.html](http://www.tigr.org/tdb/bacends/mouse/bac_end_intro.html)  
 Plate: 148 row: E column: 23

Seq primer: SP6  
 Class: BAC ends.  
 FEATURES source  
 1. .480  
 /organism="Mus musculus"  
 /strain="C57BL/6J"  
 /db\_xref="taxon:10990"  
 /clone="RPCI-23-18E23"  
 /clone\_id="RPCI-23"  
 /sex="female"  
 /lab\_host="DH10B"  
 /note="Organ: kidney/Brain; Vector: pBACE3.6; Site\_1:  
 EcoRI; Site\_2: EcoRI; Female C57BL/6J mouse kidney and/or  
 brain genomic DNA was isolated and partially digested  
 with a combination of EcoRI and EcoRI Methylase. Size  
 selected DNA was cloned into the pBACE3.6 vector at the  
 EcoRI sites. The ligation products were transformed into  
 DH10B electrocompetent cells (BRL Life Technologies)."  
 BASE COUNT 96 a 141 c 99 g 144 t  
 ORIGIN  
 Query Match 94.4%; Score 17; DB 17; Length 480;  
 Best Local Similarity 100.0%; Pred. No. 1.7e+03;  
 Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGAGCTTGGGAGC 17  
 Db 449 TGGAGCTTGGGAGC 465



clone\_end:T? vector\_side:left"

misc\_feature 74369\_101258 /note= "assembly\_name:Contig19"

misc\_feature 101359\_.112855 /note= "assembly\_name:Contig20"

misc\_feature 142956\_.215293 /note= "assembly\_name:Contig21"

BASE COUNT 62075 a 44444 c 42411 g 65360 t 1003 others

ORIGIN

Query Match, Best Local Similarity 100.0%; Score 18; DB 2; Length 215293; Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGAGGTGAGGAGCA 18 |||||||

Db 44171 TGGAGCTTGGGAGCA 44188 |||||||

RESULT 3 AC018191/c

DEFINITION Drosophila melanogaster, \*\*\* SEQUENCING IN PROGRESS \*\*\*, in ordered pieces.

ACCESSION AC018191

VERSION AC018191.1 GI:6553000

KEYWORDS HTGS\_PHASE2

SOURCE

ORGANISM Drosophila melanogaster.

JOURNAL Butaria; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydriidea; Drosophilidae; Drosophila; (bases 1 to 78508)

REFERENCE 1 Adams,M. and Venter,J.C.

TITLE Direct Submission

COMMENT Submitted (09-DEC-1999) Celera Genomics, 45 West Glade Drive, Rockville, MD, USA

This sequence was identified as CDM:10131377 by the submitter. For more information on this record e-mail to fly@celera.com.

\* NOTE: This is a working draft sequence.

\* This sequence will be replaced by the finished sequence as soon as it is available and the accession number will be preserved.

FEATURES source

1. 78508 /organism="Drosophila melanogaster" /gb\_xref="taxon:7227" /gb\_xref="taxon:7227"

BASE COUNT 23544 a 16827 c 16845 g 21292 t

ORIGIN

Query Match Best Local Similarity 100.0%; Score 17; DB 2; Length 78508; Matches 17; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 GGGACTTGGGGAGCA 18 |||||||

Db 72959 GGGACTTGGGGAGCA 72943 |||||||

RESULT 4 AC116512/c

DEFINITION Mus musculus clone RP24-328C19, WORKING DRAFT SEQUENCE, 9 ordered pieces.

ACCESSION AC116512

VERSION AC116512.3 GI:2135432

KEYWORDS HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP; SOURCE house mouse.

ORGANISM Mus musculus

REFERENCE 1 (bases 1 to 139324)

clone\_end:T? vector\_side:left"

MUS musculus, clone RP24-328C19

Unpublished

JOURNAL

2 (bases 1 to 139924)

REFERENCE

AUTHORS Birren,B., Nusbaum,C. and Lander,E.

JOURNAL

Unpublished

AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S., Barna,N., Barna,N., Barna,N., Bloom,T., Boguski,M., Boukhalter,B., Chazaro,B., Choepel,Y., Colangeli,M., Collins,S., Collymore,A., Cook,A., Cooke,P., Debreuilano,K., Devar,K., Diaz,J.S., Dodge,S., Faro,S., Ferreira,P., FitzHugh,W., Gage,D., Galadand,J., Gardyne,S., Ginde,S., Gord,S., Govertte,M., Graham,L., Grand-Pierre,N., Hagos,B., Horton,I., Huime,W., Iliev,I., Johnson,R., Jones,C., Kamat,A., Karatas,A., Keils,C., LaRoque,R., Lamzares,R., Landers,T., Lehoocky,J., Levine,R., Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J., Marquis,N., Matthews,C., McCarthy,M., McEwan,P., McKernan,K., Meidim,J., Meneus,L., Mihowa,T., Milenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Plunkett,J., Pukac,L., Rieback,M., Riley,R., Rose,C., Rogov,P., Raymond,C., Rettar,R., Rieback,M., Riley,R., Rise,C., Roman,J., Rosetti,M., Roy,A., Santos,R., Schauer,S., Schupback,R., Seaman,S., Selevny,P., Spencer,B., Stange-Thomann,N., Stojanovic,N., Straus,N., Subramanian,A., Talama,J., Testafyre,S., Theodore,J., Topham,K., Travers,M., Travas,J., Trigilio,J., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.

REFERENCE

AUTHORS

JOURNAL

Submitted (28-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 139924)

REFERENCE

AUTHORS

JOURNAL

Submitted (28-MAR-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

3 (bases 1 to 139924)

REFERENCE

AUTHORS

JOURNAL

Submitted (12-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

On Jul 12, 2002 this sequence version replaced 91:21700663.

All repeats were identified using RepeatMasker: Smit,A.P.A. & Green,P. (1996-1997) <http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/MIT Center for Genome Research

Center project name: L25512

Center clone name: 328\_C\_19

----- Summary Statistics

Sequencing vector: Plasmid; n/a: 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: phrap; version 0.960731

Consensus quality: 13716 bases at least 040

Consensus quality: 13835 bases at least 030

Consensus quality: 138779 bases at least 020

Insert size: 13000; agarose-fp

Insert size: 139124; sum-of-contigs

Quality coverage: 8.1 in Q20 bases; agarose-fp

Copyright (c) 1993 - 2003 Compugen Ltd.

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## OM nucleic - nucleic search, using sw model

Run on:

May 24, 2003, 09:39:57 ; Search time 705.6 Seconds

(without alignments)

Million cell updates/sec

Title: US-09-935-464-39

Perfect score: 21

Sequence: 1 gataacccccggcttcta-tgaag 21

Scoring table: IDBNTITY\_NUC

GapOp 10.0 , Gapext 1.0

Searched:

204640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters:

4109280

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:\*

1: gb\_gb:\*

2: gb\_htg:\*

3: gb\_in:\*

4: gb\_om:\*

5: gb\_ov:\*

6: gb\_ph:\*

7: gb\_ph:\*

8: gb\_Pl:\*

9: gb\_Pr:\*

10: gb\_ro:\*

11: gb\_sts:\*

12: gb\_sy:\*

13: gb\_un:\*

14: gb\_vl:\*

15: em\_ba:\*

16: em\_fun:\*

17: em\_hum:\*

18: em\_in:\*

19: em\_mu:\*

20: em\_om:\*

21: em\_or:\*

22: em\_ov:\*

23: em\_pat:\*

24: em\_ph:\*

25: em\_Pl:\*

26: em\_ro:\*

27: em\_sts:\*

28: em\_vl:\*

29: em\_htg\_bum:\*

30: em\_htg\_ivn:\*

31: em\_htg\_other:\*

32: em\_htg\_mus:\*

33: em\_htg\_pln:\*

34: em\_htg\_rnd:\*

35: em\_htg\_men:\*

36: em\_htg\_vrt:\*

Result No.	Score	Query Match Length	DB ID	Description
1	21	100.0	2447 6 AX399682	AX399682 Sequence
2	21	100.0	2464 9 AF428261	AF428261 Homo sapi
3	21	100.0	2474 9 BC032787	BC032787 Homo sapi
4	19.4	92.4	1738 9 HS272L16	AL045688 Human gen
5	19.4	92.4	2612 9 AK095713	AK095713 Homo sapi
6	19.4	92.4	157875 9 HS272L16	AL023754 Human DNA
7	17.4	82.9	52478 2 AC01972	AL672172 Mouse DNA
8	17.4	82.9	167353 2 AC01972	AC103972 Mus muscu
9	17.4	82.9	12273 2 AL045441	AL845441 Mus muscu
10	17.4	82.9	218502 2 AC029564	AC029564 Mus muscu
11	17.4	82.9	240536 10 AC025910	AC025910 Mus muscu
12	17	81.0	459 6 AX33172	AX43172 Sequence
13	16.8	80.0	80049 2 AC103129	AC103129 Rattus no
14	16.8	80.0	109882 2 AC111359	AC111359 Rattus no
15	16.8	80.0	161529 2 AC096081	AC096081 Rattus no
16	16.8	80.0	177639 2 AC130761	AC130761 Rattus no
17	16.8	80.0	337261 2 AC096079	AC096079 Rattus no
18	16.8	80.0	179251 2 11 G73213	AJ3213 cosus Micr
19	16.2	77.1	947 9 HS4323617	AJ323617 Homo sapi
20	16.2	77.1	1013 10 DB6557	DB6557 Rattus norv
21	16.2	77.1	1032 10 AB023027	AB023027 MUS muscu
22	16.2	77.1	1332 10 DB6556	DB6556 Rattus norv
23	16.2	77.1	1402 10 RATPRKI	L24407 Rattus norv
24	16.2	77.1	1439 10 RATCAMPKA	L26388 Rattus norv
25	16.2	77.1	1554 10 AF18184	AF18184 Mus muscu
26	16.2	77.1	1618 10 AB004267	AB004267 Rattus no
27	16.2	77.1	2470 3 LM729564	AJ295564 Leishmani
28	16.2	77.1	2416 10 BC021840	BC021840 Mus muscu
29	16.2	77.1	2427 10 AF428262	AF428262 Mus muscu
30	16.2	77.1	36068 3 AC012051	U91743 Leishmania
31	16.2	77.1	36068 3 AC009602	AC012051 Leishmani
32	16.2	77.1	37778 3 AC009602	AC009602 Leishmani
33	16.2	77.1	42070 3 S_pombe C	AL035529 S_pombe C
34	16.2	77.1	89354 9 AC008957	AC008957 Homo sapi
35	16.2	77.1	115626 9 AC010631	AC010631 Homo sapi
36	16.2	77.1	153033 2 AC044895	AC112475 Rattus no
37	16.2	77.1	153452 2 AC122530	AC044895 Homo sapi
38	16.2	77.1	154334 2 AC122530	AC122530 Mus muscu
39	16.2	77.1	155937 2 AC016341	AC016341 Homo sapi
40	16.2	77.1	156940 9 AC007444	AC007444 Homo sapi
41	16.2	77.1	160025 2 AC120619	AC120619 Rattus no
42	16.2	77.1	163046 9 AP000839	AP000839 Homo sapi
43	16.2	77.1	163054 2 AC12087	AC12087 Rattus no
44	16.2	77.1	165229 2 AC095327	AC095327 Rattus no
45	16.2	77.1	166668 2 AC126913	AC126913 Rattus no

## ALIGNMENTS

Pred. No. is the number of results predicted by chance to have a

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
Patient: WO 022947-A 3 28-MAR-2002;

Bukreyeva; Metzsoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.

DeLane, A.D. and Yegnanathan, T.

Cancer associated protein kinases and their uses

KINETEK PHARMACEUTICALS INC (CA); UNIT BRITISH COLUMBIA (CA)

FEATURES source

1. .2447 /organism="Homo sapiens" /protein\_id="CAN3\_521\_1"

/db\_xref="taxon:9606" /note="Unnamed protein product"

/codon\_Start=1 /protein\_id="CAN3\_521\_1"

/db\_xref="GI:21335456" /translation="MGRKEEDCCSSKKKTTNKRKFIFMEVLGGATESEVFVQKLIGKLFALKCKKSPAFRDSLNEAVLKRIKHNNTLEIDYESTTHYLQMLSGGEFDLTGRRVTEKADSLIQVQIQLQVLSKQVHENGIVHDLRPNLILPENSKI

MITDGLSKNQNGIMSTAGCTATGAGAVLQVAKVYHENGIVHDLRPNLILPENSKI

FYEETESKLEFEKKEKEYTETSPWDDIESKADFTCHLERDNERTCERKHPWIDGTAHLDFIPSVLQIOQNFKFNRQANAAVHNMKHMHSQVRPEVNE

NRPPEQASSISRPSSPEITTEAPVLDHVVALPAUOLPCOHRGPAPGRSNCLVNGSLHASSIVPHQSLSAGPGCCSSCCLNIGNKGKSSYCSSEPTLKKANKKQFKSEVMPVKAQGSSHCRAQGQVCLIM"

BASE COUNT ORIGIN 590 a 707 c 604 g 546 t

Query Match Best Local Similarity 100.0%; Score 21; DB 5; Length 2447; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY |||||||GATACCCCGCTCTATGAG 21

Db 713 GATACCCCGCTCTATGAG 733

RESULT 2

LOCUS AF428261 2464 bp mRNA linear PRI 07-NOV-2001

DEFINITION Homo sapiens calcium/calmodulin-dependent protein kinase I gamma (CAMK1) mRNA, complete cds.

ACCESSION AF428261.1 GI:16755791

VERSION 1

KEYWORDS ORGANISM Homo sapiens

REFERENCE Bukay-Yoata, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo. Schutte,B.C., Bjork,B.C., Coppage,K.B., Malik,M.I., Gregory,S.G., Scott,D.J., Brentzelli,L.M., Watnabe,Y., Dixon,M.J. and Murray,J.C.

A preliminary gene map for the Van der Woude syndrome critical region derived from 900 kb of genomic sequence at 1q32-q41

JOURNAL Genome Res. 10 (1), 81-94 (2000)

PUBMED 10213118

REFERENCE 1 (bases 1 to 2464)

2 (bases 1 to 2464)

AUTHORS Bjork,B.C., Watnabe,Y., Murray,J.C. and Schutte,B.C.

TITLE Characterization of the human ortholog of rat Cam Kinase I gamma (CamK1) at 1q32-q41

JOURNAL Unpublished

REFERENCE 3 (bases 1 to 2464)

AUTHORS Bjork,B.C., Watnabe,Y., Murray,J.C. and Schutte,B.C.

TITLE Direct Submission

JOURNAL Submitted (05-OCT-2001) Genetics, University of Iowa, 140 EMRB, Iowa City, IA 52242, USA

FEATURES source

1. .2464 /organism="Homo sapiens"

/gb\_xref="taxon:9606" /map="1q32-q41"

1. .2464 /gene="CAMK1G" /map="1q32-q41"

2. .1503 /gene="CAMK1G" /map="1q32-q41"

3. .1503 /product="CAMK1G" /map="1q32-q41"

/product="calcium/calmodulin-dependent protein kinase I

gamma" /protein\_id="AAL28100\_1"

/protein\_id="AAL28100\_1"

/db\_xref="GI:16755792" /translation="MGRKEEDCCSSKKKTTNKRKFIFMEVLGGATESEVFVQKLIGKLFALKCKKSPAFRDSLNEAVLKRIKHNNTLEIDYESTTHYLQMLSGGEFDLTGRRVTEKADSLIQVQIQLQVLSKQVHENGIVHDLRPNLILPENSKI

MITDGLSKNQNGIMSTAGCTATGAGAVLQVAKVYHENGIVHDLRPNLILPENSKI

FYEETESKLEFEKKEKEYTETSPWDDIESKADFTCHLERDNERTCERKHPWIDGTAHLDFIPSVLQIOQNFKFNRQANAAVHNMKHMHSQVRPEVNE

NRPPEQASSISRPSSPEITTEAPVLDHVVALPAUOLPCOHRGPAPGRSNCLVNGSLHASSIVPHQSLSAGPGCCSSCCLNIGNKGKSSYCSSEPTLKKANKKQFKSEVMPVKAQGSSHCRAQGQVCLIM"

BASE COUNT ORIGIN 609 a 708 c 602 g 545 t

Query Match Best Local Similarity 100.0%; Score 21; DB 9; Length 2464; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY |||||||GATACCCCGCTCTATGAG 21

Db 716 GATACCCCGCTCTATGAG 736

RESULT 3

LOCUS BC032787 2474 bp mRNA linear PRI 27-JUN-2002

DEFINITION Homo sapiens, calcium/calmodulin-dependent protein kinase IG, clone MGC:44894 IMAGE:5179957, mRNA, complete cds.

ACCESSION BC032787

VERSION BC032787.1 GI:21619664

KEYWORDS MGC.

ORGANISM Homo sapiens

REFERENCE Bokaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS 1 (bases 1 to 2474)

TITLE Direct Submission

JOURNAL Submitted (06-JUN-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC project URL: http://mgc.nci.nih.gov

COMMENT Contact: MGC help desk Email: chapbs@mail.nih.gov

Tissue procurement: Life Technologies, Inc.

CDNA library preparation: Life Technologies, Inc.

CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LILN)

DNA Sequencing by: National Institutes of Health Intramural Sequencing Center (NIISC), Gaithersburg, Maryland;

Web Site: http://www.nisc.nih.gov/

Contact: nisc.mucanbri.nih.gov/

Akhter,N., Ayale,K., Beckstrom-Sternberg,S.M., Benjamin,B., Blakesley,R.W., Bouffard,G.G., Breen,K., Brinkley,C., Brooks,S., Dietrich,N.L., Granite,S., Guan,X., Gupta,J., Haghjhi,P., Hansen,N., Ho,S.-L., Karins,E., Kwong,P., Laric,P., Lekas,P.R., Maduro,O.L., Mastilo,C., Maskeri,B., Mastrian,S.D., McCloskey,J.C., McDowell,J.J., Pearson,R., Staartvoop,S., Thomas,P.J., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A., Wetherby,K.D., Wiggins,L., Young,A., Zhang,L.-H. and Green,E.D.

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LILN att: http://image.llnl.gov

Series: IRAK Plate: 68 Row: k Column: 17

This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 14196444.

Location/Qualifiers

FEATURES source

1. .2474 /organism="Homo sapiens"



PR 19-OCT-2000; 2000US-0693036.  
 PR 29-NOV-2000; 2000US-0727344.  
 XX  
 PA (KINE-) KINERIK PHARM INC.  
 PA (UBR-) UNIV BRITISH COLUMBIA.  
 XX  
 PI Yoganathan T, Delaney AD;  
 XX DR WPI; 2002-394145/42.  
 DR P-PSDB; AAE22764.  
 XX PR Diagnosing cancer, comprises determining the upregulation of expression of a nucleic acid sequence encoding a protein kinase or upregulation of PT expression of the protein kinase, in the cancer -  
 XX PS Claim 16; Page 62-64; 87pp; English.  
 CC The invention relates to a method for screening biologically active agent CC that modulates cancer associated protein kinase function. The invention CC also relates to a method for diagnosing cancer comprising determining the upregulation of expression of a nucleic acid sequence encoding a protein kinase. The method is useful for diagnosing cancer. A protein kinase is useful for screening biological agents that modulate cancer associated protein kinase function. Downregulating the activity of protein kinase is useful for inhibiting the growth of a cancer cell, e.g. liver or colon cancer. A nucleic acid encoding protein kinase is useful to screen biopsy derived tumours and inflammatory samples such as arthritic synovium, for amplified DNA in the or increased expression of corresponding mRNA or protein and is also useful to detect differences in expression levels such as molecular weight, amino acid and nucleotide sequences between the two cells. The present sequence is human calmodulin kinase CAMK-XI gene located on chromosome 1q32.1-32.3.  
 XX SQ Sequence 2447 BP; 590 A; 707 C; 604 G; 546 T; 0 other;  
 Query Match 100.0%; Score 21; DB 24; Length 2447;  
 Best Local Similarity 100.0%; Pred. No. 0.22; Matches 21; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
 CC QY 1 GATACCCCGGCTCATGAG 21  
 ||||||| ||||| ||||| ||||| |||||  
 Db 713 GATACCCCGGCTCATGAG 733  
 RESULT 2  
 AAI60703  
 ID AAI60703 standard; cDNA; 1956 BP.  
 XX AC AAI60703;  
 XX DT 22-OCT-2001 (first entry)  
 XX DE Human polynucleotide SEQ ID NO 4692.  
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager syndrome; chemotactic; chemoattractant; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.  
 XX OS Homo sapiens.  
 XX PN WO200153312-A1.  
 XX PD 26-JUL-2001.  
 XX PF 26-DEC-2000; 2000WO-US34263.  
 XX PR 21-JAN-2000; 2000US-0486725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-059042.  
 PR 19-JUL-2000; 2000US-0620312.  
 OS 03-APR-2000; 2000US-0654450.  
 XX 24-SEP-2000; 2000US-0662191.

PR Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries  
 XX PS Claim 1; SEQ ID NO 4692; 1007pp; English.  
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAI8642-AAI42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis, and inflammation, leukaemias and C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed specification.  
 XX SQ Sequence 1956 BP; 474 A; 559 C; 508 G; 414 T; 1 other;  
 Query Match 92.4%; Score 19.4; DB 22; Length 1956;  
 Best Local Similarity 95.2%; Pred. No. 1.5;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 CC QY 1 GATACCCCGGCTCATGAG 21  
 ||||||| ||||| ||||| ||||| |||||  
 Db 708 GATACCCCGGCTCATGAG 728  
 RESULT 3  
 AAI58917  
 ID AAI58917 standard; cDNA; 2165 BP.  
 XX AC AAI58917;  
 XX DT 22-OCT-2001 (first entry)  
 XX DE Human polynucleotide SEQ ID NO 1120.  
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager syndrome; chemotactic; chemoattractant; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.  
 XX OS Homo sapiens.  
 XX PN WO200153312-A1.  
 XX PD 26-JUL-2001.  
 XX PF 26-DEC-2000; 2000WO-US34263.  
 XX PR 21-JAN-2000; 2000US-0486725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-059042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-APR-2000; 2000US-0654450.  
 PR 24-SEP-2000; 2000US-0662191.  
 PR 09-JUL-2000; 2000US-0598042.

XX  
 PA (HYSE-) HYSEQ INC.  
 PA (HYSE-) HYSEQ INC.  
 XX PI Tang YY, Liu C, Asundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D; Wang J, Wehrman T, Xu C, Xue AJ, Yang Y, Zhang J;  
 PI Zhao QA, Zhou P, Goodrich R, Dzmaniac RT;  
 XX DR WPI; 2001-442253/47.  
 DR P-PSDB; AAM41547.  
 XX PR Novel nucleic acids and polypeptides, useful for treating disorders such as central nervous system injuries  
 XX PS Claim 1; SEQ ID NO 4692; 1007pp; English.  
 CC The invention relates to human nucleic acids (AAI57798-AAI61369) and the encoded polypeptides (AAI8642-AAI42213) with nootropic, immunosuppressant and cytostatic activity. The polynucleotides are useful in gene therapy. A composition containing a polypeptide or polynucleotide of the invention may be used to treat diseases of the peripheral nervous system, such as peripheral nervous injuries, peripheral neuropathy and localised neuropathies and central nervous system diseases, such as Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic lateral sclerosis, and Shy-Drager Syndrome. Other uses include the utilisation of the activities such as: immune system suppression, Activin/inhibin activity, chemotactic/chemokinetic activity, haemostatic and thrombolytic activity, cancer diagnosis and therapy, drug screening, assays for receptor activity, arthritis, and inflammation, leukaemias and C.N.S disorders.  
 CC Note: The sequence data for this patent did not form part of the printed specification.  
 XX SQ Sequence 1956 BP; 474 A; 559 C; 508 G; 414 T; 1 other;  
 Query Match 92.4%; Score 19.4; DB 22; Length 1956;  
 Best Local Similarity 95.2%; Pred. No. 1.5;  
 Matches 20; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 CC QY 1 GATACCCCGGCTCATGAG 21  
 ||||||| ||||| ||||| ||||| |||||  
 Db 708 GATACCCCGGCTCATGAG 728  
 RESULT 2  
 AAI60703  
 ID AAI60703 standard; cDNA; 1956 BP.  
 XX AC AAI60703;  
 XX DT 22-OCT-2001 (first entry)  
 XX DE Human polynucleotide SEQ ID NO 4692.  
 XX Human; nootropic; immunosuppressant; cytostatic; gene therapy; cancer; peripheral nervous system; neuropathy; central nervous system; CNS; Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic; amyotrophic lateral sclerosis; Shy-Drager syndrome; chemotactic; chemoattractant; thrombolytic; drug screening; arthritis; inflammation; leukaemia; ss.  
 XX OS Homo sapiens.  
 XX PN WO200153312-A1.  
 XX PD 26-JUL-2001.  
 XX PF 26-DEC-2000; 2000WO-US34263.  
 XX PR 21-JAN-2000; 2000US-0486725.  
 PR 25-APR-2000; 2000US-0552317.  
 PR 09-JUL-2000; 2000US-059042.  
 PR 19-JUL-2000; 2000US-0620312.  
 PR 03-APR-2000; 2000US-0654450.  
 PR 24-SEP-2000; 2000US-0662191.  
 PR 09-JUL-2000; 2000US-0598042.